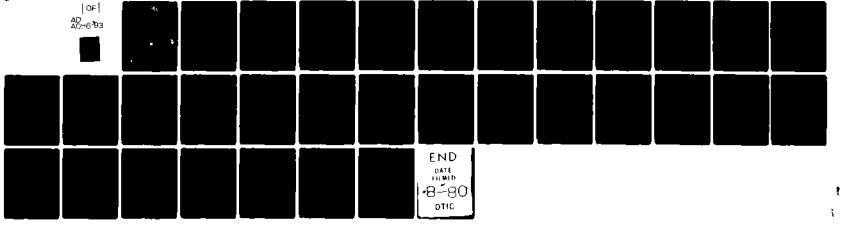
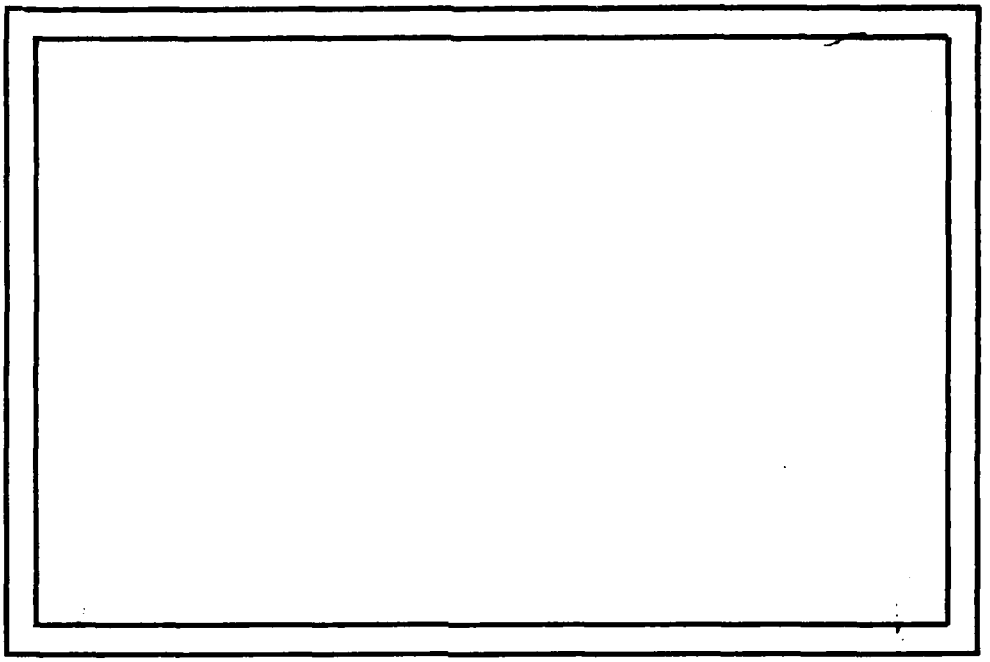
AD-A086 193 MARYLAND UNIV COLLEGE PARK DEPT OF COMPUTER SCIENCE F/G 12/1
COMPUTABLE ERROR BOUNDS FOR AGGREGATED MARKOV CHAINS,(U)
MAY 80 G W STEWART N00014-76-C-0391
UNCLASSIFIED TR-901 NL



LEVEL II

6

AD A086193








# UNIVERSITY OF MARYLAND
# COMPUTER SCIENCE CENTER

COLLEGE PARK, MARYLAND
20742



80 6 5 015

(9) Technical Report (14) TR-901

(11) May 1980

(12) 33

# (6) Computable Error Bounds for Aggregated Markov Chains

(10) G. W. Stewart



## Abstract

This paper describes a method for computing the steady state probability vector of a nearly completely decomposable Markov chain. The method is closely related to one proposed by Simon and Ando [7] and developed by Courtois [2,3]. However, the method described here does not require the determination of a completely decomposable stochastic approximation to the transition matrix and hence it is applicable to matrices other than stochastic. An error analysis of the procedure is given which results in effectively computable error bounds.



*Department of Computer Science and Institute for Physical Science and Technology, The University of Maryland. This work was supported in part by the Office of Naval Research under Contract No (15) N00014-76-C-0391

# Computable Error Bounds for Aggregated Markov Chains

G. W. Stewart

## 1. Introduction

This paper will be concerned with techniques for treating a discrete, finite Markov chain whose matrix of transition probabilities can, after a suitable renumbering of the states, be written in the form

$$A = \begin{bmatrix} A_{11} & E_{12} & \cdots & E_{1\ell} \\ E_{21} & A_{22} & \cdots & E_{2\ell} \\ \vdots & \vdots & & \vdots \\ E_{\ell 1} & E_{\ell 2} & \cdots & A_{\ell\ell} \end{bmatrix}, \tag{1.1}$$

where the matrices $E_{ij}$ are small. The matrix $A$ is nonnegative and stochastic; i.e.

$$A\underset{\sim}{1} = \underset{\sim}{1} ,$$

so that the vector $\underset{\sim}{1}$ consisting of all ones is a right eigenvector of $A$ corresponding to the eigenvalue one. If, in addition, $A$ is irreducible*, the eigenvalue one is simple and there is a unique, normalized, positive left eigenvector $y$ corresponding to the eigenvalue one (in the irreducible case we shall call the eigenvalue the Perone root). If $A$ is acyclic and $y$ is normalized so that $\underset{\sim}{1}^T y = 1$, then $y$ is the vector of steady state probabilities for the chain. One of

* For terminology, see [14].

the chief computational problems associated with Markov chains is the determination of the vector $y$.

Chains with transition matrices of the form (1.1) are said to be nearly completely decomposible. They arise naturally as models of systems whose states can be clustered into aggregates that are loosely connected to one another. They were first studied by Simon and Ando [7], who had applications to economic systems in mind. A recent monograph by Courtois [3] contains a history of the subject and extensive applications in the computer sciences.

The usual computational procedure goes as follows. The off-diagonal blocks $E_{ij}$ are amalgamated into the diagonal blocks $A_{ii}$ to produce a block diagonal approximation $A^*$ to $A$ that has the form

$$A^* = \text{diag}(A^*_{11}, A^*_{22}, \ldots, A^*_{\ell\ell}) \quad .$$

This decomposition is done in such a way that each block $A^*_{ii}$ is stochastic and irreducible. The steady state vectors $y_i$ of the $A^*_{ii}$ are then computed and the steady state vector of the original system approximated in the form

$$y \cong \begin{bmatrix} \nu_1 y^*_1 \\ \nu_2 y^*_2 \\ \vdots \\ \nu_\ell y^*_\ell \end{bmatrix} \tag{1.2}$$

The quantities $\nu_i$ are calculated as the components of an eigenvector of a matrix of order $\ell$ whose elements may be easily calculated from

the vectors $y_i^*$ and the original matrix A. The computational advantages of this method are obvious, since it reduces the solution of a large eigenvalue problem to that of several potentially much smaller ones.

The purpose of this paper is to resolve two difficulties with the method as it is currently practiced. The first concerns the determination of the approximating decomposed matrix $A^*$, a process frequently refered to by the unfortunate term "aggregation".* There are infinitely many ways to incorporate the off-diagonal blocks of A into the diagonal blocks in order to get an approximation $A^*$. In some instances this flexibility may be useful. For example [13], in certain highly structured systems it is possible to determine the diagonal blocks $A_{ii}^*$ so that the eigenvectors $y_i^*$ are exactly proportional to the corresponding pieces of y [cf. (1.2)]. In general, however, the indeterminacy of $A^*$ is a nuisance; some choices of $A^*$ may be better than others, but without further information there is no way of knowing. In particular, the derivation of any general error bound for the approximation (1.2) must necessarily entail the assumption that the worst choice has been made. In this paper a new method is proposed that does not require intermediate approximations but works directly with the original matrix A.

The second problem treated here is that of showing how reasonably sharp error bounds may be computed. Courtois [2, 3] has given an error analysis for the procedure sketched above, which shows in part how it

* By all natural usage, "aggregation" should refer to the determination of which states are to be clustered together.

behaves as the off-diagonal blocks become small. However, the analysis is not suitable for computing error bounds for two reasons. First, the analysis is asymptotic in the size of the off-diagonal blocks, and it is not shown how small the blocks must be for it to be approximately correct. Second, the analysis assumes that all the matrices involved have complete systems of eigenvectors. Although it is unlikely that any given problem will fail to have this property, it is not at all unlikely that it will be near a problem that does, in which case an analysis based on eigenvector expansions will give unrealistic results owing to the ill condition of the matrix of normalized eigenvectors.

The techniques developed in this paper are not restricted to stochastic matrices; rather they can be applied to find the dominant eigenvalue of almost any matrix of the form (1.1). What is required is that the dominant eigenvalues of the $A_{ii}$ be simple and have sufficiently well-conditioned eigenvectors and that the $E_{ij}$ be sufficiently small. If $A$ is stochastic, these conditions are likely to be satisfied; but as will be seen, the computational techniques test the conditions directly, without reference to the properties of $A$. In particular, if one of the $A_{ii}$ has the form (1.1), its dominant eigenvectors can be found independently of $A$ by the method described in this paper. This observations has important consequences for the process of multi-level aggregation described by Courtois [3].

This paper is organized as follows. Sections 2 and 3 lay the theoretical foundations for the techniques to follow; Section 2 describes the deflation of a simple eigenvalue, and Section 3 reviews perturbation theory

for invariant subspaces. In Section 4 the technique is sketched broadly, and in Section 5 it is justified in detail by the derivation of effectively computable error bounds. In Section 6 the practical techniques from numerical analysis required to implement the method are discussed. The paper concludes with a numerical example.

Many of the results of this paper will be cast in terms of vector and matrix norms. The symbol $\|\cdot\|$ will denote either the Euclidean vector norm defined by

$$\|x\|^2 = x^T x$$

or the spectral matrix norm defined by

$$\|A\| = \max_{\|x\|=1} \|Ax\| .$$

The symbol $\|\cdot\|_F$ will denote the Frobenius matrix norm defined by

$$\|A\|_F^2 = \sum_{i,j} a_{ij}^2 .$$

Note that for any vector $x$,

$$\|x\| = \|x\|_F .$$

For more on these norms see [8].

It is important not to expect too much of error bounds cast in terms of norms. In the first place, repeated use of inequalities such as the triangle inequality tends to make them pessimistic. In the second place,

such a bound can be difficult to interpret in terms of the components of the vector thus bounded. For example, if $\|e\| \leq \varepsilon$, then any component of $e$ can be as large as $\varepsilon$. But other things being equal, it is more likely that each component is of order $\varepsilon/\sqrt{n}$.

In cases where an error bound is unsatisfactory, it may be necessary to calculate an error estimate, in which an attempt is made to approximate the error vector itself. For many problems this can be done, although frequently a heavy computational price must be paid. Moreover, once an error estimate has been calculated, it is hard to resist the temptation to use it to improve the putative solution, which will set off another round of error bounding. To feel the force of this temptation, the reader is invited to consider the table of estimates given on page 187 of [3].

## 2. The constructive theory of a simple eigenvalue.

In this section are collected a number of results about a simple eigenvalue and its eigenvectors which can be found in one form or another scattered throughout the literature. The results follow from a constructive reduction of $A$ to block diagonal form by means of rather simple similarity transformations.

Let $A$ be a matrix of order $n$ with a real simple eigenvalue $\beta$ corresponding to the eigenvector $x$. Since $x$ is nonzero, it may be normalized so that $\|x\| = 1$. Let the columns of $n\times(n-1)$ matrix $Y$ form an orthonormal basis for the subspace orthogonal to $x$; i.e.

$$Y^TY = I \tag{2.1}$$

and

$$Y^T x = 0 \quad . \tag{2.2}$$

This implies that the matrix $(x \;\; Y)$ is orthogonal.

Consider the similarity transformation

$$(x \;\; Y)^T A (x \;\; Y) = \begin{bmatrix} x^T Ax & x^T AY \\ Y^T Ax & Y^T AY \end{bmatrix} \tag{2.3}$$

$$= \begin{bmatrix} \beta x^T x & x^T AY \\ \beta Y^T x & Y^T AY \end{bmatrix} .$$

It follows from (2.2) and the fact that $x^T x = 1$ that

$$(x \;\; Y)^T A (x \;\; Y) = \begin{bmatrix} \beta & g^T \\ 0 & C \end{bmatrix} ,$$

where

$$g^T = x^T AY \quad ,$$

$$C = Y^T AY \quad .$$

The matrix $C$ has for its eigenvalues the eigenvalues of $A$ other than $\beta$, hence $C - \beta I$ is nonsingular.

Consider now the further similarity transformation

$$\begin{bmatrix} 1 & -q^T \\ 0 & I \end{bmatrix} \begin{bmatrix} \beta & g^T \\ 0 & C \end{bmatrix} \begin{bmatrix} 1 & q^T \\ 0 & I \end{bmatrix} = \begin{bmatrix} \beta & \beta q^T + g^T - q^T C \\ 0 & C \end{bmatrix} \quad . \tag{2.4}$$

Since $C - \beta I$ is nonsingular, $q$ may be chosen to satisfy

$$(C - \beta I)^T q = g \quad ,$$

from which it follows that the row vector in the upper right of (2.4) is zero. Thus the two similarity transformations (2.3) and (2.4) reduce $A$ to the block diagonal form $\mathrm{diag}(\beta, B)$.

The composite similarity transformation that reduces $A$ can be found by multiplying the two transformations (2.3) and (2.4). Specifically, set

$$(x \quad X) = (x \quad Y) \begin{bmatrix} 1 & q^T \\ 0 & I \end{bmatrix} = (x \quad Y + xq^T)$$

and

$$(y \quad Y) = (x \quad Y) \begin{bmatrix} 1 & 0 \\ -q & I \end{bmatrix} = (x - Yq \quad Y) \quad . \tag{2.5}$$

Then

$$(x \quad X)^{-1} = \begin{bmatrix} y^T \\ Y \end{bmatrix} \tag{2.6}$$

and

$$\begin{bmatrix} y^T \\ Y \end{bmatrix} A \, (x \quad X) = \begin{bmatrix} y^T Ax & 0 \\ 0 & Y^T AX \end{bmatrix} = \begin{bmatrix} \beta & 0 \\ 0 & C \end{bmatrix} \tag{2.7}$$

A number of important facts can be read from this reduction. In the first place $y$ is the left eigenvector of $A$ corresponding to $\beta$. Since from (2.5), $y^T x = 1$, it follows that $y$ is not orthogonal to $x$.

Since from (2.5)

$$y = x - Yq \quad ,$$

an alternate expression for $q$ follows from (2.1) and (2.2) :

$$q = - Y^T y \quad .$$

Moreover,

$$\| y \|^2 = 1 + \| q \|^2 \quad .$$

Similarly,

$$\| X \|^2 = 1 + \| q \|^2 \quad .$$

All these results may be summarised in the following theorem.

Theorem 2.1. Let $\beta$ be a simple eigenvalue of a matrix $A$ of order $n$, and let the corresponding eigenvector $x$ be normalized so that

$$1. \quad \| x \| = 1 \quad .$$

Let $y$ be the left eigenvector corresponding to $\beta$. Then $y$ is not orthogonal to $x$ and may be normalized so that

$$2. \quad y^T x = 1 \quad .$$

Moreover there are $nx(n-1)$ matrices $X$ and $Y$ such that

$$3. \quad Y^T Y = Y^T X = I$$

$$4. \qquad y^T X = x^T Y = 0$$

$$5. \qquad (x \;\; X)^{-1} = (y \;\; Y)^T$$

$$6. \qquad (y \;\; Y)^T A (x \;\; X) = \begin{bmatrix} \beta & 0 \\ 0 & C \end{bmatrix} ,$$

where

$$7. \qquad C = Y^T AX = Y^T AY \quad .$$

The eigenvalues of the matrix $C$ are the eigenvalues of $A$ other than $\beta$. If $q$ is defined by either of the expressions

$$8. \qquad q = (C - \beta I)^{-T} Y^T A^T x \quad ,$$

$$9. \qquad q = -Y^T y \quad ,$$

then

$$10. \qquad X = Y + xq^T \quad ,$$

$$11. \qquad y = x - Yq \quad ,$$

$$12. \qquad \| X \|^2 = \| y \|^2 = 1 + \| q \|^2 \quad .$$

## 3. Perturbation theory

In this section the following problem will be addressed: given a matrix $A$ partitioned in the form

$$A = \begin{bmatrix} B & G \\ H & C \end{bmatrix} ,$$

find a matrix $U$ as near as possible to the identity such that the transformed matrix $\bar{A} = UAU^{-1}$ has the form

$$\bar{A} = \begin{bmatrix} \bar{B} & 0 \\ \bar{H} & \bar{C} \end{bmatrix} .$$

The importance of this problem lies in the following observation. If $v$ is a left eigenvector of $\tilde{B}$, then $(v^T, 0)$ will be a left eigenvector of $\bar{A}$, and $(v^T, 0)U$ will be a left eigenvector of $A$. Since

$$\frac{\| (v^T, 0)U - (v^T, 0) \|}{\| (v^T, 0) \|} \leq \| I - U \| ,$$

the vector $(v^T, 0)$ will be a good approximate left eigenvector of $A$ in proportion as $U$ is near the identity matrix.

This problem has been treated in [9], and the following is a summary of the results required in this paper. The reader is referred to the reference for proofs.

The problem will have a solution only if the eigenvalues of $B$ and $C$ are separated and $G$ is sufficiently small. Unfortunately, the minimum of the distances between the eigenvalues of $B$ and $C$ is too crude a measure of separation to give satisfactory bounds. Instead the measure

$$\delta(B, C) = \inf_{\|P\|_F = 1} \| BP - PC \|_F$$

will be used. The properties of $\delta(B, C)$ are summarized in the following theorem.

Theorem 3.1. The number $\beta$ is zero if and only if $B$ and $C$ have an eigenvalue in common. Moreover,

1. $\delta(B, C) \le \min\left\{|\beta - \gamma| : \begin{matrix}\beta \text{ an eigenvalue of } B \\ \gamma \text{ an eigenvalue of } C\end{matrix}\right\}$,

2. $\delta(B+E, C+F) \ge \delta(B, C) - \|E\| - \|F\|$ ,

3. $\delta[\mathrm{diag}(B_1, \ldots, B_p), \mathrm{diag}(C_1, \ldots, C_q)] = \min\{\delta(B_i, C_j): i=1, \ldots, p;\ j=1, \ldots, q\}$ ,

4. $\delta(\beta, C) = \|(\beta I - C)^{-1}\|^{-1}$ .

Properties 2, 3, and 4 in the theorem are particularly important in computational practice. Property 2 says that small changes in $B$ and says $C$ make equally small changes in $\delta(B, C)$, a property not shared by the minimum distance between the eigenvalues of $B$ and $C$. Property 3 shows how $\delta$ for a block diagonal matrix can be found from the $\delta$'s between the blocks. Finally, property 4 gives an explicit expression for $\delta$ when one of the matrices is a scalar. These properties will be used extensively in the derivation of the error bounds in §5.

The solution to the problem posed at the beginning of this section requires that $U$ be chosen in a specific form. Specifically, $U$ will be written

$$U = \begin{bmatrix} I & -P \\ P^T & I \end{bmatrix} \begin{bmatrix} (I + PP^T)^{-1/2} & 0 \\ 0 & (I + P^TP)^{-1/2} \end{bmatrix} . \tag{3.1}$$

Here $(I + PP^T)^{-1/2}$ is the inverse of the unique positive definite square

root of the positive definite matrix $I + PP^T$ -- similarly for $(I + P^TP)^{-1/2}$. It is easily verified that $U$ is orthogonal; i.e. $U^TU = I$ or $U^{-1} = U^T$. Thus the problem becomes that of determining $P$ so that

$$U^T \begin{bmatrix} B & G \\ H & C \end{bmatrix} U = \begin{bmatrix} \bar{B} & 0 \\ \bar{H} & \bar{C} \end{bmatrix} . \tag{3.2}$$

Conditions under which this can be done are contained in the following theorem.

Theorem 3.2. In the notation introduced above, let

$$\gamma \geq \|G\|_F , \qquad \eta \geq \|H\|_F ,$$
$$\delta \leq \delta(B, C) . \tag{3.3}$$

If

$$\tau \equiv \frac{\eta\gamma}{\delta^2} < \frac{1}{4} , \tag{3.4}$$

then there exists a unique matrix $P$ satisfying

$$\|P\|_F \leq \frac{\gamma}{\delta} \frac{1 + \sqrt{1-4\tau}}{1 - 2\tau + \sqrt{1-4\tau}} < 2\frac{\gamma}{\delta} \tag{3.5}$$

such that $U$ defined by (3.1) satisfies (3.2). Moreover,

$$\bar{B} = (I + PP^T)^{-1/2} (B + PH) (I + PP^T)^{1/2} . \tag{3.6}$$

The distance of $U$ from the identity matrix is roughly $\|P\|_F$ .

The bound (3.5) shows that this distance depends linearly on $\|G\|_F$ and inversely on $\delta(B, C)$. In other words, the bound becomes smaller as G becomes small and larger as the spectra of B and C approach one another; however, the bounds can be considerably worse than a naive inspection of the spectra would indicate (cf. 1 in Theorem 3.1).

The expression for $\overline{B}$ is particularly interesting. When $\|H\| = O(\|G\|)$, as it will in §5, both $PP^T$ and PH are $O(\|G\|^2)$ and as G approaches zero the eigenvalues of B approximate those at the original matrix up to terms of $O(\|G\|^2)$. This quadratic behavior will prove critical in deriving workable error bounds.

## 4. The approximation algorithm

In this section an algorithm for approximating the dominant eigenvector of a matrix of the form (1.1) will be described. It is assumed that the diagonal elements $A_{ii}$ are all irreducible. In order to keep the exposition simple, the algorithm will be described for a 3x3 partitioning, i.e. $\ell = 3$. The general case is a trivial extension.

For each i, let $\beta_i > 0$ be the Perone root of $A_{ii}$ and let $x_i > 0$ be its corresponding right eigenvector. Since $\beta_i$ is simple, $A_i$ has a decomposition of the form described in Theorem 2.1, viz.

$$(y_i, Y_i)^T A (x_i, X_i) = \begin{bmatrix} \beta_i & 0 \\ 0 & C_i \end{bmatrix} .$$

It then follows from Theorem 2.1 that the inverse of the matrix

$$X = \begin{bmatrix} x_1 & 0 & 0 & X_1 & 0 & 0 \\ 0 & x_2 & 0 & 0 & X_2 & 0 \\ 0 & 0 & x_3 & 0 & 0 & X_3 \end{bmatrix}$$

is

$$Y^T = \begin{bmatrix} y_1^T & 0 & 0 \\ 0 & y_2^T & 0 \\ 0 & 0 & y_3^T \\ Y_1^T & 0 & 0 \\ 0 & Y_2^T & 0 \\ 0 & 0 & Y_3^T \end{bmatrix} .$$

Consider the matrix

$$(4.1) \qquad Y^T AX = \begin{bmatrix} \beta_1 & \phi_{12} & \phi_{13} & 0 & g_{12}^T & g_{13}^T \\ \phi_{21} & \beta_2 & \phi_{23} & g_{21}^T & 0 & g_{22}^T \\ \phi_{31} & \phi_{32} & \beta_3 & g_{31}^T & g_{32}^T & 0 \\ 0 & h_{12} & h_{13} & C_1 & F_{12} & F_{13} \\ h_{21} & 0 & h_{23} & F_{21} & C_2 & F_{23} \\ h_{31} & h_{32} & 0 & F_{31} & F_{32} & C_3 \end{bmatrix} \equiv \begin{bmatrix} B & G \\ H & C \end{bmatrix} ,$$

where

$$\phi_{ij} = y_i^T E_{ij} x_j , \qquad g_{ij}^T = y_i^T E_{ij} X_j ,$$

$$h_{ij} = Y_i^T E_{ij} x_j , \qquad F_{ij} = Y_i^T E_{ij} X_j$$

Because $y_i > 0$ and $x_j > 0$, it follows that $\phi_{ij} = 0$ if and only if $E_{ij} = 0$. Hence, B is irreducible, since A is. Let $\rho$ be the Perone root of B with right eigenvector u and left eigenvector

$$v^T = (\nu_1, \nu_2, \nu_3) \quad .$$

The approximation to the left eigenvector y of A is then given by

$$\tilde{y} = \begin{bmatrix} \nu_1 y_1 \\ \nu_2 y_2 \\ \nu_3 y_3 \end{bmatrix} \quad . \tag{4.2}$$

This algorithm is extremely simple. All that it requires is the calculation of the left and right Perone vectors of the diagonal blocks of A, the formation of the matrix B from the $E_{ij}$, and the calculation of the left Perone vector of B. Except for the initial grouping of states to get the partition (1.1), the process is entirely deterministic, requiring no assimilation of the matrices $E_{ij}$ into the diagonal blocks $A_{ii}$.

## 5. Error bounds

In this section error bounds for the approximation (4.2) will be derived. The bounds provide a formal proof of convergence of the algorithm, as well as considerable insight into its behavior. The practical computation of the bounds will be discussed in the next section.

The approach is to use Theorem 3.2 to obtain an exact expression for $y$ in terms of a vector $\overline{v}$ that is the left eigenvector of a matrix $\overline{B}$ lying very near to $B$. A second application of Theorem 3.2 bounds $\|v - \overline{v}\|$, and hence the error in the approximation (4.2) to $y$.

For the first step, the notation of Theorem 3.2 coincides exactly with the notation of the approximation algorithm. Consequently if (3.2) is satisfied, there is a matrix $U$ of the form (3.1) that reduces $Y^TAX$ to the form (3.2). Now the eigenvalues of $\overline{B}$ are eigenvalues of $A$. Let $\overline{v}$ be the left eigenvector of $B$ corresponding to the Perone root of $A$. Then

$$y^T = (\overline{v}^T, 0)U^TY^T = \overline{v}^T(I + PP^T)^{-1/2}\,(I, P)Y^T \quad . \tag{5.1}$$

The relation between $\overline{v}$ and $v$ must now be considered. Let

$$\|P\|_F < \pi \tag{5.2}$$

be any bound, presumably obtained by an application of Theorem 3.2. As in the first part of the development of §2, extend $v$ to an orthogonal matrix $(v, V)$ such that

$$(v, V)^T B\,(v, V) = \begin{bmatrix} \rho & 0 \\ r & R \end{bmatrix} \tag{5.3}$$

and let

$$\delta_\rho \le \delta(\rho, R) \quad .$$

Let

$$(v, V)^T (B + PH) (v, V) = \begin{bmatrix} \tilde{\rho} & s \\ \tilde{r} & \tilde{R} \end{bmatrix} .$$

By Theorem 3.1

$$\delta(\rho, R) \ge \delta_\rho - 2\pi\eta \quad .$$

Moreover,

$$\|\tilde{r}\| \le \|r\| + \pi\eta \quad , \tag{5.4}$$

and

$$\|s\| \le \pi\eta \quad . \tag{5.5}$$

Hence by Theorem 3.2 if

$$\frac{\pi\eta(\|r\| + \pi\eta)}{(\delta_\rho - 2\pi\eta)^2} < \frac{1}{4}$$

there is an eigenvector

$$\tilde{v} = v + e$$

of $B + PH$ satisfying

$$|v - \tilde{v}| = |e| \le \frac{2\pi\eta}{\delta_\rho - 2\pi\eta} \tag{5.6}$$

From (3.6) we have

$$\bar{v} = \tilde{v}(I + PP^T)^{1/2} = (v + e)\ (I + PP^T)^{1/2} \quad .$$

Hence from (5.1)

$$y^T = (v + e)^T\ (I,\ P\ )Y^T =$$

$$= v^T \begin{bmatrix} y_1^T & 0 & 0 \\ 0 & y_2^T & 0 \\ 0 & 0 & y_3^T \end{bmatrix} + v^T P \begin{bmatrix} Y_1^T & 0 & 0 \\ 0 & Y_2^T & 0 \\ 0 & 0 & Y_3^T \end{bmatrix} +$$

$$+ e^T \begin{bmatrix} y_1^T & 0 & 0 \\ 0 & y_2^T & 0 \\ 0 & 0 & y_3^T \end{bmatrix} + e^T P \begin{bmatrix} Y_1^T & 0 & 0 \\ 0 & Y_2^T & 0 \\ 0 & 0 & Y_3^T \end{bmatrix}$$

But the first term in this sum is the approximation (4.2) to $y$. Hence, since $|y_i| = 1$, the final bound becomes

$$(5.7) \qquad \left\| y - \begin{bmatrix} \nu_1 y_1 \\ \nu_2 y_2 \\ \nu_3 y_3 \end{bmatrix} \right\| < \pi + \frac{2\pi\eta}{\delta_\rho - 2\pi\eta} \max\{\|y_i\|\} + \frac{2\pi^2\eta}{\delta_\rho - 2\pi\eta} \quad .$$

It is important to note that in deriving the bound (5.7) it has been implicitly assumed that the Perone root of $A$ was to be found in $\overline{B}$ and that this root corresponded to the Perone root of $B$. This cannot be insured a priori without making further assumptions. Essentially what is required is that the eigenvalues of the $C_{ii}$, be sufficiently removed from the $\beta_i$'s so that subsequent perturbations cannot turn one of them

into a Perone root. Although a formal analysis is possible, it will not be given here, since in practice it is easy to see whether the largest eigenvalue of $\overline{B}$ is the Perone root.

When $A$ is stochastic, the $O(\|E\|^2)$ perturbation in passing from $B$ to $B + PH$ is critical to the analysis. This is because all the $\beta_i$ are within $O(\|E\|)$ of unity, so that a perturbation of $O(\|E\|)$ will completely scramble the eigenvectors.

For computational purposes we have scaled the $x_i$ and $y_i$ so that $\|x_i\| = 1$ and

$$y_i^T x_i = 1 \quad . \tag{5.8}$$

In fact the approximation algorithm will yield the same results for any scaling, provided only that (5.8) is satisfied. To see this, suppose that $x_i$ is replaced by $\hat{x}_i = \delta_i x_i$, where $\delta_i$ is a nonzero scaling factor. Then (5.8) requires that $y_i$ be replaced by $\hat{y}_i = \delta_i^{-1} y_i$. It is easily seen that this results in $B$ being replaced by $D^{-1}BD$, where $D = \text{diag}(\delta_1, \delta_2, \delta_3,)$. Consequently the left eigenvector of $D^{-1}BD$ is $v^T D$ and the approximation to the left eigenvector of the entire system is $(\delta_1 \nu_1 \hat{y}_1^T, \delta_2 \nu_2 \hat{y}_2^T, \delta_3 \nu_3 \hat{y}_3^T) = (\nu_1 y_1^T, \nu_2 y_2^T, \nu_3 y_3^T)$, which is the same as (4.2).

For stochastic matrices there is a natural scaling of the $y_i$ that leads to a beautiful interpretation of the approximation process. Specifically, let $y_i$ be scaled so that

$$\underset{\sim}{1}^T y_i = 1 \quad , \tag{5.9}$$

i.e. so that $y_i$ can be interpreted as a vector of probabilities. Write

$$x_i = \underset{\sim}{1} + p_i \quad .$$

If $x_i$ is given the scaling implied by (5.8), then

$$y_i^T p_i = y_i^T x_i - y_i^T \underset{\sim}{1} = 1 - 1 = 0 \quad ,$$

and

$$y_i^T A_{ii} p_i = \beta_i y_1^T p_i = 0 \quad . \tag{5.10}$$

Moreover, since $A_{ii}$ is within $0(\|E\|)$ of a stochastic matrix, it follows from Theorems 3.1 and 3.2 that if $\delta(\beta_i, C_i)$ is large enough the vector $p_i$ will satisfy

$$p_i = 0(\|E\|) \quad . \tag{5.11}$$

It will now be shown that, up to terms of $0(\|E\|^2)$, the matrix $B$ in (4.1) is stochastic. Consider the first row sum

$$\begin{aligned} \beta_1 + \phi_{12} + \phi_{13} &= y_1^T A_{11} x_1 + y_1^T E_{12} x_2 + y_1^T E_{13} x_3 \\ &= y_1^T (A_{11} \underset{\sim}{1} + E_{12} \underset{\sim}{1} + E_{13} \underset{\sim}{1}) + y_1^T A_{11} p_1 + y_1^T (E_{12} p_2 + E_{13} p_3) \\ &= y_1^T (A_{11} \underset{\sim}{1} + E_{12} \underset{\sim}{1} + E_{13} 1) + 0(\|E\|^2) \end{aligned}$$

by (5.10) and (5.11). Because $A$ is stochastic $A_{11}\underset{\sim}{1} + E_{12}\underset{\sim}{1} + E_{13}\underset{\sim}{1} = \underset{\sim}{1}$. Hence

$$\beta_1 + \phi_{12} + \phi_{13} = y^T \underset{\sim}{1} + 0(\|E\|^2) = 1 + 0(\|E\|^2) \quad ,$$

by (5.9). Thus the first row sum of B is within $O(\|E\|^2)$ of one, and so are the other row sums.

The nearly stochastic matrix B -- or rather $\bar{B}$ which differs from B by $O(\|E\|^2)$ -- controls the long term behavior of the Markov chain. To see this note that by an application of Theorem 2.1 the matrix (4.1) can be reduced to the form $\text{diag}(\bar{B}, \bar{C})$. Now the behavior of the Markov chain is controlled by the behavior of the powers $A^k$ $(k = 1,2,3,\ldots)$, and this behavior can be determined by examining the behavior of the powers of $\text{diag}(\bar{B}, \bar{C})$.

Specifically, $\text{diag}(\bar{B}, \bar{C})^k = \text{diag}(\bar{B}^k, \bar{C}^k)$. Since the eigenvalues of $\bar{C}$ are less than those of $\bar{B}$, the powers will approach $\text{diag}(\bar{B}^k, 0)$. Since $\bar{B}$ has a dominant eigenvalue of one, $\bar{B}^k$ will tend more slowly to the matrix $\bar{v}\,\bar{w}^T$, where $\bar{v}$ and $\bar{w}$ are the left and right eigenvectors corresponding to one. In terms of the original Markov chain, if the state vector $y^{(k)}$ is written in the form

$$y^{(k)} = \begin{bmatrix} \nu_1^{(k)} y_1^{(k)} \\ \nu_2^{(k)} y_2^{(k)} \\ \nu_3^{(k)} y_3^{(k)} \end{bmatrix},$$

where $\underset{\sim}{1}^T y_i^{(k)} = 1$, then the $y_i^{(k)}$ will converge swiftly as $\bar{C}^k \to 0$ and the $v_i^{(k)}$ will converge more leisurely as $\bar{B}^k$ approaches its limit. This justifies calling B the long term transition matrix of the chain.

Of course this double limit behavior of nearly completely decomposable chains has been remarked by numerous researchers, beginning with Simon

and Ando [7]; the approach taken here merely makes explicit the factors that control the rates of convergence. Although the matter will not be persued in this paper, it should be possible to obtain numerical convergence rates from an analysis of the behavior of the powers of the matrices B and $C_i$.

## 6. Practical details

The details of the implementation of the approximation algorithm and the computation of the bounds will depend on the sizes of the matrices involved. Three classes of matrices may be distinguished

1. Matrices that can be stored as an array in the high speed memory of a computer. Typically, an upper bound for the order of such matrices ranges from fifty to five hundred, depending on the computer.

2. Matrices that cannot be stored in array form but whose structure permits the efficient solution of a system of linear equations with the matrix elements as coefficients. Examples of such matrices are band matrices and "sparse" matrices [4]. Their orders can be very large.

3. Matrices that are so large that the only thing one can do with them is to form their product with a vector.

Each of these classes will be discussed in turn.

If the matrices $A_{ii}$ lie in the first class, the appropriate procedure is to use the QR algorithm to reduce $A_{ii}$ to quasi-triangular form. Specifically, there exists software [11] to compute an orthogonal matrix $(x_i, Y_i)$, such that

$$(x_i, Y_i)^T A_{ii} (x_i, Y_i) = \begin{bmatrix} \beta_i & g_i^T \\ 0 & C_i \end{bmatrix}$$

where $C_i$ is quasi-triangular, i.e. block upper triangular with 1x1 and 2x2 blocks on its diagonal. Because $C_i$ is quasi-triangular it is extremely cheap to solve linear systems involving $C_i - \beta_i I$, which means that it is practical to compute $q$ in part 8 of Theorem 2.1, from which $y_i$ can be calculated.

The next step is to compute $B$. Since this requires a pass over the matrices $E_{ij}$, this is also a logical time to compute the bounds $\eta$ and $\gamma$ in (3.3). In very large problems it is unlikely that there will be storage to contain the matrices $X_i$ and $Y_i$, so that $H$ and $G$ cannot be computed explicitly. However, bounds may be obtained by the following procedure. For each $E_{ij}$, compute $E_{ij}^T x_j$ and $y_i^T E_{ij}$. Then set

$$\eta^2 = \sum_{i,j} \| Y_i^T \|^2 \| E_{ij} x_j \|^2 = \sum_{i,j} \| E_{ij} x_j \|^2 \tag{6.1}$$

$$\gamma^2 = \sum_{i,j} \| y_i^T E_{ij} \|^2 \| X_i \|^2 = \sum_{i,j} \| y_i^T E_{ij} \|^2 \| y_i \|^2 \quad . \tag{6.2}$$

A bound $\phi_c$ on the off-diagonal elements of $C$ will be needed later.

It can be calculated at this point in the form

$$(6.3) \qquad \phi_c = \sum_{i,j} \|Y_i^T\| \, \|E_{ij}\| \, \|X_j\| = \sum_{i,j} \|E_{ij}\| \, \|X_j\|$$

The application of the perturbation theorem requires a lower bound $\delta$ on $\delta(B, C)$. Because $C_i$ is quasi-triangular, $\|(\beta - C)^{-1}\|$ can be cheaply estimated by a variant of the inverse power method [1,6]. Set $\beta^+ = \max\{\beta_i\}$ and $\beta^- = \min\{\beta_i\}$. Then set

$$(6.4) \qquad \delta = \min \|(\beta^+ I - C_i)^{-1}\|^{-1} - (\beta^+ - \beta^-) - \phi_b - \phi_c \quad ,$$

where $\phi_b$ is any upper bound on the norm of the off-diagonal part of $B$. It can easily be shown by repeated appeals to Theorem 3.1 that $\delta$ is indeed a lower bound for $\delta(B, C)$.

Having computed $\eta$, $\gamma$, and $\delta$, the bound $\pi$ in (5.2) can be computed from (3.5). The matrix $B$ may now be analysed in the same manner as the $A_{ii}$ [cf. (5.3)] to get a bound on $\delta(\rho, R)$. At this point the bounds (5.4) and (5.5) may also be computed. Finally, the bound (5.7) the accuracy of the approximation may be computed .

For systems of large sparse matrices, it is possible to compute right and left eigenvectors $x_i$ and $y_i$ by means of the inverse power method [ 8 ]. However, it will not generally be possible to maintain the matrices $X_i$, $Y_i$, and $C_i$ in the high-speed memory of the computer in question. Fortunately the matrix $(x_i, Y_i)$ can be written as a Householder transformation in the form

$$(x_i, Y_i) = I - 2w_i w_i^T \quad ,$$

where $w_i$ is a vector of norm unity that can be determined from $x_i$ alone (for details see [8]). If $x_i^T y_i = 1$, then it follows from Theorem 2.1 that $\|X_i\| = \|y_i\|$. Hence $\eta$, $\gamma$, and $\phi_c$ may be estimated as in (6.1), (6.2), and (6.3). Finally, since from item 7 in Theorem 2.1, it follows that $C_i = Y_i^T A_{ii} Y_i$, the separation $\delta(B, C)$ may be bounded as described above, but this time using the technique of implicit deflation [12] to solve systems involving $Y_i^T A Y_i - \beta I$. The rest of the bounding process proceeds as above.

For matrices of the third class, i.e. matrices for which only matrix-vector multiplication is possible, the eigenvectors $x_i$ and $y_i$ must be computed by the power method or its variants [5, 10]. The numbers $\eta$ and $\gamma$ may be computed as described above. Unfortunately, there is no way of computing a lower bound on $\delta(B, C)$ since there is no way of solving linear systems involving $C_i = Y_i^T A Y_i$. The best that can be done is to compute an approximate upper bound that may be a reasonable estimate. From item 8 in Theorem 2.1 we have

$$\|q\| \leq \|(C_i - \beta_i I)^{-T}\| \, \|Y_i^T A_{ii}^T x_i\| ,$$

and from item 12,

$$\|(C_i - \beta_i I)^{-1}\|^{-1} \leq \frac{\|Y_i^T A_{ii}^T x_i\|}{(\|y_i\|^2 - 1)^{1/2}} .$$

Hence, in analogy with (6.4), it may be hoped that the number

$$\delta = \min \frac{\| y_i^T A_{ii}^T x_i \|}{( \| y_i \|^2 - 1)^{1/2}} - (\beta^+ - \beta^-) - \phi_b - \phi_c$$

will not differ too much from $\delta(B, C)$.

## 6. A numerical example.

In this section the techniques of this paper will be applied to a matrix analysed previously by Courtois. The matrix with its partitioning is given in Table 1.*

Table 2 gives the details of the calculation of the approximation $\tilde{y}$ in (4.2). The vector is compared with the true vector $y$, which has been scaled so that $\| y - \tilde{y} \|$ is a minimum. Table 3 gives the details of the computation of the error bound (5.7).

The error bound is good enough for practical purposes, even though it is an order of magnitude bigger than the actual error. This overestimate is in part due to the repeated use of norm inequalities in the derivation of the bound. It is also due to the fact that Theorem 3.2 bounds $\| P \|_F$, whereas it is clear that the smaller spectral norm could be used in the derivation of (5.7). As was pointed out in the introduction, a pessimistic view of the error is inevitable when on attempts to bound it rather than estimate it.

## 7. Acknowledgement

The author is greatful to Jon Agre for performing the numerical calculations reported in §6 and especially for his interest and encouragement.

---

* The matrix, as reported in [2,3] is not stochastic; the sixth row does not sum to one. Since the property of being stochastic is not necessary to the procedure described in this paper, the matrix is left as reported.

Table 1

The Matrix A and its Partitioning [2,3]

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 0.85000 | 0.00000 | 0.14900 | 0.00090 | 0.00000 | 0.00005 | 0.00000 | 0.00005 |
| 0.10000 | 0.65000 | 0.24900 | 0.00000 | 0.00090 | 0.00005 | 0.00000 | 0.00005 |
| 0.10000 | 0.80000 | 0.09960 | 0.00030 | 0.00000 | 0.00000 | 0.00010 | 0.00000 |
| 0.00000 | 0.00040 | 0.00000 | 0.70000 | 0.29950 | 0.00000 | 0.00010 | 0.00000 |
| 0.00050 | 0.00000 | 0.00040 | 0.39900 | 0.60000 | 0.00010 | 0.00000 | 0.00000 |
| 0.00000 | 0.00050 | 0.00000 | 0.00000 | 0.00005 | 0.60000 | 0.24990 | 0.15000 |
| 0.00003 | 0.00000 | 0.00003 | 0.00004 | 0.00000 | 0.10000 | 0.80000 | 0.09990 |
| 0.00000 | 0.00005 | 0.00000 | 0.00000 | 0.00005 | 0.19990 | 0.25000 | 0.55000 |

Table 2

Computation of $\tilde{y}$ (4.2)

| $\beta_i$ | $x_i$ | $y_i$ | | | |
|---|---|---|---|---|---|
| 0,9991 | | 0.9993 | | 0.9993 | |
| 0.5772 | 0.6954 | 0.7074 | 0.8080 | 0.5774 | 0.4170 |
| 0.5773 | 0.7218 | 0.7069 | 0.6061 | 0.5774 | 0.9624 |
| 0.5776 | 0.3150 | | | 0.5774 | 0.3527 |

B

| | | |
|---|---|---|
| 0.9991 | 0.0010 | 0.0001 |
| 0.0005 | 0.9993 | 0.0001 |
| 0.0002 | 0.0001 | 0.9999 |

$v^T$

| | | |
|---|---|---|
| 0.4433 | 0.6130 | 0.6539 |

| $\tilde{y}$ | y |
|---|---|
| 0.308281 | 0.307971 |
| 0.320022 | 0.320321 |
| 0.139653 | 0.139726 |
| 0.495361 | 0.495323 |
| 0.371565 | 0.371616 |
| 0.272669 | 0.272734 |
| 0.629330 | 0.629282 |
| 0.230652 | 0.230659 |

$$\frac{\|\tilde{y} - y\|}{4.5 \cdot 10^{-4}}$$

Table 3

Computation of the Error Bound

Norms of Off-diagonal Matrices

$\eta = \|H\|_F = 4.54 \cdot 10^{-4}$

$\gamma = \|G\|_F = 2.14 \cdot 10^{-4}$

$\phi_b = 9.76 \cdot 10^{-4}$

$\phi_c = 9.29 \cdot 10^{-4}$

Computation of $\delta$ (6.4) and $\pi$ (3.5)

$\beta_+ = 0.9999$ $\beta_- = 0.9991$

$\|(\beta_+ - C_{ii})^{-1}\|^{-1} = 0.2352, 0.6992, 0.4487$

$\delta = 0.2325$

$\tau = 0.0013$ [from (3.4)]

$\pi = 9.22 \cdot 10^{-4}$

Final error bound

$\delta_\rho = 2.81 \cdot 10^{-4}$

$\|r\| = 3.095$ [cf (5.3)]

$\tau = 0.0017$

$\|e\| \leq 2.99 \cdot 10^{-3}$ [cf (5.6)]

$\|\tilde{y} - y\| \leq 4.23 \cdot 10^{-3}$

## References


1. A. K. Cline, C. B. Moler, G. W. Stewart, and J. H. Wilkinson, An estimate for the condition number, SIAM J. Numer. Anal. 16(1979) 368-375.

2. P. J. Courtois, Error analysis in nearly-completely decomposable stochastic systems, Econometrica 43(1975) 691-709.

3. _____, Decomposability: Queueing and Computer System Applications, Academic Press, New York (1977).

4. Iain Duff and G. W. Stewart, Sparse Matrix Proceedings 1978, SIAM, Philadelphia (1979)

5. A. Jennings and W. J. Stewart, Simultaneous iteration for partial eigensolution of real matrices, J. Inst. Math. Appl. 15(1975) 351-362.

6. D. P. O'Leary, G. W. Stewart, J. S. Vandergraft, Estimating the largest eigenvalue of a positive definite matrix, Math. Comp. 33(1979) 1289-1292.

7. H. A. Simon and A. Ando, Aggregation of variables in dynamic systems, Econometrica 29(1961) 111-138.

8. G. W. Stewart, Introduction to Matrix Computations, Academic Press, New York (1974).

9. _____, Error and Perturbation Bounds for Subspaces Associated with certain eigenvalue problems, SIAM Rev. 15(1973) 727-764.

10. _____, Simultaneous iteration for computing invariant subspaces of non-Hermitian matrices. Numer. Math.25(1976) 123-136.

11. _____, HQR3 and EXCHNG: Fortran subroutines for calculating and ordering the eigenvalues of a real upper Hessenberg matrix, ACM Trans. Math. Software 3 (1976) 275-280.

12. _____, On the implicit deflation of nearly singular systems of linear equations, University of Maryland, Computer Science Technical Report Series TR-437(1976).

13. H. Vantilborgh, Exact aggregation in exponential queueing networks, J. Assoc. Comput. Mach. 25(1978) 620-629.

14. R. S. Varga, Matrix Iterative Analysis, Prentice-Hall, Englewood Cliffs, New Jersey (1962).